Multivariate isotonic time series regression

Michael H. Neumann

Friedrich-Schiller-Universität Jena

(with **Konstantinos Fokianos, Anne Leucht**)

Motivation: Estimation of integer-valued time series

$(Y_t)_{t \in \mathbb{N}}$ such that

$$Y_t \mid Y_{t-1}, Y_{t-2}, \ldots \sim \text{Poisson}(\lambda_t),$$

Motivation: Estimation of integer-valued time series

$(Y_t)_{t\in\mathbb{N}}$ such that

$$Y_t \mid Y_{t-1}, Y_{t-2}, \ldots \sim \text{Poisson}(\lambda_t),$$

where e.g.

$$\lambda_t = f(Y_{t-1}, \lambda_{t-1}, t/n, Z_t),$$

Motivation: Estimation of integer-valued time series

$(Y_t)_{t \in \mathbb{N}}$ such that

$$Y_t \mid Y_{t-1}, Y_{t-2}, \ldots \sim \text{Poisson}(\lambda_t),$$

where e.g.

$$\lambda_t = f(Y_{t-1}, \lambda_{t-1}, t/n, Z_t),$$

i.e., conditional expectation depends on

- lagged variable(s), lagged intensity/ies, trend, exogenous variable(s)

Motivation: Estimation of integer-valued time series

$(Y_t)_{t\in\mathbb{N}}$ such that

$$Y_t \mid Y_{t-1}, Y_{t-2}, \ldots \sim \text{Poisson}(\lambda_t),$$

where e.g.

$$\lambda_t = f(Y_{t-1}, \lambda_{t-1}, t/n, Z_t),$$

i.e., conditional expectation depends on

- lagged variable(s), lagged intensity/ies, trend, exogenous variable(s)

nonparametric (time series) regression:

$$Y_t = f(Y_{t-1}, \lambda_{t-1}, t/n, Z_t) + \varepsilon_t, \qquad E(\varepsilon_t \mid Y_{t-1}, \ldots) = 0 \quad a.s.$$

Motivation: Estimation of integer-valued time series

$(Y_t)_{t \in \mathbb{N}}$ such that

$$Y_t \mid Y_{t-1}, Y_{t-2}, \ldots \sim \text{Poisson}(\lambda_t),$$

where e.g.

$$\lambda_t = f(Y_{t-1}, \lambda_{t-1}, t/n, Z_t),$$

i.e., conditional expectation depends on

- lagged variable(s), lagged intensity/ies, trend, exogenous variable(s)

nonparametric (time series) regression:

$$Y_t = f(Y_{t-1}, \lambda_{t-1}, t/n, Z_t) + \varepsilon_t, \qquad E(\varepsilon_t \mid Y_{t-1}, \ldots) = 0 \quad a.s.$$

particular challenge: explanatory variables not "regularly" distributed
$\Longrightarrow$ suitability of standard nonparametric methods (kernel,...) not clear

Motivation: Estimation of integer-valued time series

On the other hand...
... linear models, e.g. integer-valued AR(p):

$$\lambda_t = \theta_0 + \theta_1 Y_{t-1} + \cdots + \theta_p Y_{t-p}$$

Motivation: Estimation of integer-valued time series

On the other hand...
... linear models, e.g. integer-valued AR(p):

$$\lambda_t = \theta_0 + \theta_1 Y_{t-1} + \cdots + \theta_p Y_{t-p}$$

Natural condition: $\theta_0 > 0$, $\theta_1, \ldots, \theta_p \geq 0$

Motivation: Estimation of integer-valued time series

On the other hand...

... linear models, e.g. integer-valued AR(p):

$$\lambda_t = \theta_0 + \theta_1 Y_{t-1} + \cdots + \theta_p Y_{t-p}$$

Natural condition: $\theta_0 > 0$, $\theta_1, \ldots, \theta_p \geq 0$

$\Longrightarrow f\colon \mathbb{R}^p \to \mathbb{R}$ isotonic (non-decreasing in each component)

Isotonic regression: Classical least squares estimator

$$Y_t = f(I_t) + \varepsilon_t, \quad t = 1, \ldots, n,$$

- I_t d-dimensional information variable
- $E(\varepsilon_t \mid I_t) = 0$ a.s.
- $f\colon \mathbb{R}^d \to \mathbb{R}$ isotonic

Isotonic regression: Classical least squares estimator

$$Y_t = f(I_t) + \varepsilon_t, \quad t = 1, \ldots, n,$$

- I_t d-dimensional information variable
- $E(\varepsilon_t \mid I_t) = 0$ a.s.
- $f\colon \mathbb{R}^d \to \mathbb{R}$ isotonic

Isotonic least squares estimator:

$$\widetilde{f}_n \in \arg \min_{g \text{ isotonic}} \sum_{t=1}^{n} (Y_t - g(I_t))^2$$

Isotonic regression: Classical least squares estimator

Advantages of $\widetilde{f}_n$:

- no bandwidth choice necessary
- $d = 1$: irregular distribution of I_t doesn't harm,
  $\int |\widetilde{f}_n(x) - f(x)| \, dP^{I_t}(x)$ converges with optimal rate $n^{-1/3}$

Isotonic regression: Classical least squares estimator

Advantages of $\widetilde{f}_n$:

- no bandwidth choice necessary
- $d = 1$: irregular distribution of I_t doesn't harm,
  $\int |\widetilde{f}_n(x) - f(x)| \, dP^{I_t}(x)$ converges with optimal rate $n^{-1/3}$

Appropriateness of isotonicity condition?

- many parametric models for integer-valued time series produce isotonic conditional mean functions
- applications in many areas:
  biology, medicine, statistics, psychology, genetics
  (see. e.g. Luss, Rosset & Shahar (*Ann. Statist.*, 2012))

Classical isotonic least squares estimator: Details

Alternative representation of $\widetilde{f}_n$:

If $x \in \{X_1, \ldots, X_n\}$, then (Brunk, 1955)

$$\begin{aligned}\widetilde{f}_n(x) &= \max_{U:\, x \in U} \min_{L:\, x \in L} \mathrm{Av}_Y(L \cap U) \\ &= \min_{L:\, x \in L} \max_{U:\, x \in U} \mathrm{Av}_Y(L \cap U),\end{aligned}$$

Classical isotonic least squares estimator: Details

Alternative representation of $\widetilde{f}_n$:

If $x \in \{X_1, \ldots, X_n\}$, then (Brunk, 1955)

$$\begin{aligned}\widetilde{f}_n(x) &= \max_{U:\, x \in U} \min_{L:\, x \in L} \mathrm{Av}_Y(L \cap U) \\ &= \min_{L:\, x \in L} \max_{U:\, x \in U} \mathrm{Av}_Y(L \cap U),\end{aligned}$$

where

$$\mathrm{Av}_Y(S) = \frac{\sum_{t=1}^{n} Y_t \, \mathbb{1}(I_t \in S)}{\#\{t \leq n: \quad I_t \in S\}},$$

Classical isotonic least squares estimator: Details

Alternative representation of $\widetilde{f}_n$:
If $x \in \{X_1, \dots, X_n\}$, then (Brunk, 1955)

$$\widetilde{f}_n(x) = \max_{U:\, x \in U} \min_{L:\, x \in L} \mathrm{Av}_Y(L \cap U) = \min_{L:\, x \in L} \max_{U:\, x \in U} \mathrm{Av}_Y(L \cap U),$$

where

$$\mathrm{Av}_Y(S) = \frac{\sum_{t=1}^n Y_t \, \mathbb{1}(I_t \in S)}{\#\{t \le n: \quad I_t \in S\}},$$

- L lower set
  $(y \in L,\ z \preceq y \Rightarrow z \in L)$
- U upper set
  $(y \in U,\ z \succeq y \Rightarrow z \in U)$
- $z \preceq y$ means $z_i \le y_i$
  $\forall i = 1, \dots, d$

Classical isotonic least squares estimator: Details

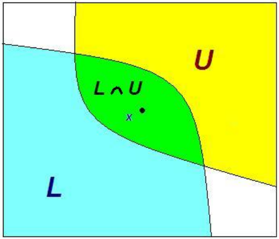
Alternative representation of $\widetilde{f}_n$:
If $x \in \{X_1, \ldots, X_n\}$, then (Brunk, 1955)

$$\widetilde{f}_n(x) = \max_{U:\, x \in U} \min_{L:\, x \in L} \mathrm{Av}_Y(L \cap U) = \min_{L:\, x \in L} \max_{U:\, x \in U} \mathrm{Av}_Y(L \cap U),$$

where

$$\mathrm{Av}_Y(S) = \frac{\sum_{t=1}^n Y_t \, \mathbb{1}(I_t \in S)}{\#\{t \le n: \quad I_t \in S\}},$$

- L lower set
  ($y \in L,\ z \le y \Rightarrow z \in L$)
- U upper set
  ($y \in U,\ z \ge y \Rightarrow z \in U$)
- $z \le y$ means $z_i \le y_i$
  $\forall i = 1, \ldots, d$



A modified estimator

- $d = 1$: $\widetilde{f}_n$ can be easily analyzed, rate of convergence $n^{-1/3}$

A modified estimator

- $d = 1$: $\widetilde{f}_n$ can be easily analyzed, rate of convergence $n^{-1/3}$
- $d > 1$:
  - Hanson, Pledger & Wright (1973): $\widetilde{f}_n$ consistent at continuity points

A modified estimator

- $d = 1$: $\widetilde{f}_n$ can be easily analyzed, rate of convergence $n^{-1/3}$
- $d > 1$:
  - Hanson, Pledger & Wright (1973): $\widetilde{f}_n$ consistent at continuity points
  - problem: huge number of lower and upper sets
    (metric and bracketing entropies very large; see e.g. Gao & Wellner (2007) and Wu, Meyer & Opsomer (2015))

A modified estimator

- $d = 1$: $\widetilde{f}_n$ can be easily analyzed, rate of convergence $n^{-1/3}$
- $d > 1$:
  - Hanson, Pledger & Wright (1973): $\widetilde{f}_n$ consistent at continuity points
  - problem: huge number of lower and upper sets (metric and bracketing entropies very large; see e.g. Gao & Wellner (2007) and Wu, Meyer & Opsomer (2015))
  - Gao & Wellner (2007): In density estimation, suboptimal rate of convergence for the (histogram-based) MLE if $d \geq 2$ ...

A modified estimator

- $d = 1$: $\widetilde{f}_n$ can be easily analyzed, rate of convergence $n^{-1/3}$
- $d > 1$:
  - Hanson, Pledger & Wright (1973): $\widetilde{f}_n$ consistent at continuity points
  - problem: huge number of lower and upper sets (metric and bracketing entropies very large; see e.g. Gao & Wellner (2007) and Wu, Meyer & Opsomer (2015))
  - Gao & Wellner (2007): In density estimation, suboptimal rate of convergence for the (histogram-based) MLE if $d \geq 2$ ...

Alternative estimators:

$$\overline{f}_n(\nu) = \max_{a:\, a \leq \nu} \min_{b:\, b \geq \nu} \mathrm{Av}_{\boldsymbol{Y}}([\![a, b]\!]),$$

$$\overline{\overline{f}}_n(\nu) = \min_{b:\, b \geq \nu} \max_{a:\, a \leq \nu} \mathrm{Av}_{\boldsymbol{Y}}([\![a, b]\!])$$

where $[\![a, b]\!] = [a_1, b_1] \times \cdots \times [a_d, b_d]$

A modified estimator

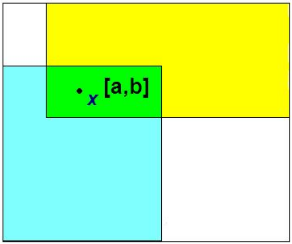
- $d = 1$: $\widetilde{f}_n$ can be easily analyzed, rate of convergence $n^{-1/3}$
- $d > 1$:
  - Hanson, Pledger & Wright (1973): $\widetilde{f}_n$ consistent at continuity points
  - problem: huge number of lower and upper sets (metric and bracketing entropies very large; see e.g. Gao & Wellner (2007) and Wu, Meyer & Opsomer (2015))
  - Gao & Wellner (2007): In density estimation, suboptimal rate of convergence for the (histogram-based) MLE if $d \geq 2$ ...

Alternative estimators:

$$\overline{f}_n(\nu) = \max_{a:\, a \leq \nu} \min_{b:\, b \geq \nu} \mathrm{Av}_{\boldsymbol{Y}}([\![a, b]\!]),$$

$$\overline{\overline{f}}_n(\nu) = \min_{b:\, b \geq \nu} \max_{a:\, a \leq \nu} \mathrm{Av}_{\boldsymbol{Y}}([\![a, b]\!])$$

where $[\![a, b]\!] = [a_1, b_1] \times \cdots \times [a_d, b_d]$



A modified estimator

If $x \notin \{I_1, \ldots, I_n\}$, take care that only averages over rectangles $[\![a, b]\!]$ with $[\![a, b]\!] \cap \{I_1, \ldots, I_n\} \neq \varnothing$ are taken ...

A modified estimator

If $x \notin \{I_1, \ldots, I_n\}$, take care that only averages over rectangles $[\![a, b]\!]$ with $[\![a, b]\!] \cap \{I_1, \ldots, I_n\} \neq \varnothing$ are taken ...

Then:

- $\overline{f}_n$, $\overline{\overline{f}}_n$ isotonic (by construction)
- converge with desired rate, no bandwidth choice required

A modified estimator

If $x \notin \{I_1, \ldots, I_n\}$, take care that only averages over rectangles $[\![a, b]\!]$ with $[\![a, b]\!] \cap \{I_1, \ldots, I_n\} \neq \varnothing$ are taken ...

Then:

- $\overline{f}_n$, $\overline{\overline{f}}_n$ isotonic (by construction)
- converge with desired rate, no bandwidth choice required

Proposed estimator:

$$\widehat{f}_n(x) = \frac{\overline{f}_n(x) + \overline{\overline{f}}_n(x)}{2}$$

Main result: What could be expected?

- $f\colon\ [0,1]^d \longrightarrow \mathbb{R}$ isotonic + bounded (+ differentiable)
  $\Rightarrow \int \sum_{i=1}^d |\partial_i f| < \infty$,
  i.e. degree of smoothness $\beta = 1$
  $\Rightarrow$ (standard) rate of convergence $n^{-\beta/(2\beta+d)} = n^{-1/(2+d)}$

Main result: What could be expected?

- $f\colon\ [0,1]^d \longrightarrow \mathbb{R}$ isotonic + bounded (+ differentiable)
  $\Rightarrow \int \sum_{i=1}^d |\partial_i f| < \infty$,
  i.e. degree of smoothness $\beta = 1$
  $\Rightarrow$ (standard) rate of convergence $n^{-\beta/(2\beta+d)} = n^{-1/(2+d)}$
- No guarantee for a good pointwise behavior!

Main result: What could be expected?

- $f\colon\ [0,1]^d \longrightarrow \mathbb{R}$ isotonic + bounded (+ differentiable)
  $\Rightarrow \int \sum_{i=1}^d |\partial_i f| < \infty$,
  i.e. degree of smoothness $\beta = 1$
  $\Rightarrow$ (standard) rate of convergence $n^{-\beta/(2\beta+d)} = n^{-1/(2+d)}$
- No guarantee for a good pointwise behavior!
- Since smoothness is measured in L_1, we consider L_1-loss, $\int |\widehat{f_n} - f| \, dP^{I_1}$

Main result (... a special case)

$$Y_t = f(I_t) + \varepsilon_t, \quad t = 1, \ldots, n,$$

- $f\colon \mathbb{R}^d \to \mathbb{R}$ isotonic, bounded

Main result (... a special case)

$$Y_t = f(I_t) + \varepsilon_t, \quad t = 1, \ldots, n,$$

- $f\colon \mathbb{R}^d \to \mathbb{R}$ isotonic, bounded
- $(I_t)_{t \in \mathbb{N}}$ stationary, I_t has a continuous distribution,
- $E(\varepsilon_t \mid I_1, \ldots, I_t, \varepsilon_1, \ldots, \varepsilon_{t-1}) = 0 \quad a.s.$,
- $E(\varepsilon_t^2 \mid I_1, \ldots, I_t, \varepsilon_1, \ldots, \varepsilon_{t-1}) \leq M < \infty \quad a.s.$

Main result (... a special case)

$$Y_t = f(I_t) + \varepsilon_t, \quad t = 1, \dots, n,$$

- $f\colon \mathbb{R}^d \to \mathbb{R}$ isotonic, bounded
- $(I_t)_{t \in \mathbb{N}}$ stationary, I_t has a continuous distribution,
- $E(\varepsilon_t \mid I_1, \dots, I_t, \varepsilon_1, \dots, \varepsilon_{t-1}) = 0 \quad a.s.,$
- $E(\varepsilon_t^2 \mid I_1, \dots, I_t, \varepsilon_1, \dots, \varepsilon_{t-1}) \le M < \infty \quad a.s.$

$$\underline{c}\, Q_1 \otimes \cdots \otimes Q_d(\cdot) \le P(I_t \in \cdot) \le \overline{c}\, Q_1 \otimes \cdots \otimes Q_d(\cdot),$$

for some $0 < \underline{c} \le \overline{c} < \infty$, $Q_1, \dots, Q_d$ probability distributions

Main result (... a special case)

$$Y_t = f(I_t) + \varepsilon_t, \quad t = 1, \ldots, n,$$

- $f\colon \mathbb{R}^d \to \mathbb{R}$ isotonic, bounded
- $(I_t)_{t \in \mathbb{N}}$ stationary, I_t has a continuous distribution,
- $E(\varepsilon_t \mid I_1, \ldots, I_t, \varepsilon_1, \ldots, \varepsilon_{t-1}) = 0 \quad a.s.,$
- $E(\varepsilon_t^2 \mid I_1, \ldots, I_t, \varepsilon_1, \ldots, \varepsilon_{t-1}) \le M < \infty \quad a.s.$

  $$\underline{c}\, Q_1 \otimes \cdots \otimes Q_d(\cdot) \le P(I_t \in \cdot) \le \overline{c}\, Q_1 \otimes \cdots \otimes Q_d(\cdot),$$

  for some $0 < \underline{c} \le \overline{c} < \infty$, $Q_1, \ldots, Q_d$ probability distributions
- $(I_t)_t$ strong $(\alpha-)$ mixing, $\sum_{r=1}^{\infty} r^{d-1/2} \alpha(r) < \infty$

Main result: A special case

- information variable I_t has a density, bounded from zero on $[0,1]^q$
- $h_n = n^{-1/(2+d)}$ (≈ optimal bandwidth)
- $D_n = [h_n, 1 - h_n]^d$

Main result: A special case

- information variable I_t has a density, bounded from zero on $[0,1]^q$
- $h_n = n^{-1/(2+d)}$ ($\approx$ optimal bandwidth)
- $D_n = [h_n, 1 - h_n]^d$

Theorem 3.1.

$$\int_{D_n} \left| \widehat{f_n}(z) - f(z) \right| dP^{I_1}(z) = O_P\left(n^{-1/(2+d)} \right).$$

Proof of Theorem 3.1

$$\int_{D_n} \left|\widehat{f}_n(z) - f(z)\right| \, dP^{I_1}(z)$$
$$= \int_{D_n} \left(\widehat{f}_n(z) - f(z)\right)_+ dP^{I_1}(z) + \int_{D_n} \left(\widehat{f}_n(z) - f(z)\right)_- dP^{I_1}(z)$$

Proof of Theorem 3.1

$$\int_{D_n} \left|\widehat{f}_n(z) - f(z)\right| \, dP^{I_1}(z)$$
$$= \int_{D_n} \left(\widehat{f}_n(z) - f(z)\right)_+ dP^{I_1}(z) + \int_{D_n} \left(\widehat{f}_n(z) - f(z)\right)_- dP^{I_1}(z)$$

For appropriate $b_n = b_n(z) \geq z$:

$$\left(\widehat{f}_n(z) - f(z)\right)_+$$
$$\leq \left(\max_{a:\, a \leq z} \mathrm{Av}_Y([\![a, b_n]\!]) - f(z)\right)_+$$

Proof of Theorem 3.1

$$\int_{D_n} \left|\widehat{f}_n(z) - f(z)\right| dP^{I_1}(z)$$
$$= \int_{D_n} \left(\widehat{f}_n(z) - f(z)\right)_+ dP^{I_1}(z) + \int_{D_n} \left(\widehat{f}_n(z) - f(z)\right)_- dP^{I_1}(z)$$

For appropriate $b_n = b_n(z) \geq z$:

$$\left(\widehat{f}_n(z) - f(z)\right)_+$$
$$\leq \left(\max_{a:\, a \leq z} \mathrm{Av}_Y([\![a, b_n]\!]) - f(z)\right)_+$$
$$\leq \left(f(b_n) - f(z)\right) + \left(\max_{a:\, a \leq z} \mathrm{Av}_\varepsilon([\![a, b_n]\!])\right)_+$$

Proof of Theorem 3.1 (cont'd)

Since $\sup_z f(z) - \inf_z f(z) < \infty$,

$$\int_{D_n} (f(b_n) - f(z)) \, dP^{I_t}(z) = O\left(n^{-1/(2+d)}\right).$$

Theorem 1 of Bickel & Wichura (1971) (on fluctuations of an empirical process...):

$$\int_{D_n} \left(\max_{a:\, a \leq z} \mathrm{Av}_\varepsilon([\![a, b_n]\!]) \right)_+ dP^{I_t}(z) = O_P\left(n^{-1/(2+d)}\right).$$

Summary

- Time series regression: (general case)

$$Y_t = f(I_t) + \varepsilon_t, \quad t = 1, \ldots, n$$

Summary

- Time series regression: (general case)

$$Y_t = f(I_t) + \varepsilon_t, \quad t = 1, \ldots, n$$

- $I_t = (X_t', Z_t')'$, $X_t \in \mathbb{N}_0^p$, Z_t q-dim., continuous
- $f\colon \mathbb{R}^{p+q} \to \mathbb{R}$ isotonic and bounded

Summary

- Time series regression: (general case)

$$Y_t = f(I_t) + \varepsilon_t, \quad t = 1, \ldots, n$$

- $I_t = (X_t', Z_t')'$, $X_t \in \mathbb{N}_0^p$, Z_t q-dim., continuous
- $f\colon \mathbb{R}^{p+q} \to \mathbb{R}$ isotonic and bounded
- $\widehat{f}_n$ (modified) nonparametric isotonic estimator
- no choice of smoothing parameter required
- rate of convergence: $n^{-1/(1+q)}$ (optimal for degree of smothness $\beta = 1$, dimension q)